不完備情報のマルコフ過程における多段決定問題の性質について (不確実・不確定性下での意思決定過程)

中井, 達

数理解析研究所講究録
第 1682 巻 2010 年 78-85

不完備情報のマルコフ過程における多段決定問題の性質について

千葉大学教育学部　中井　達 (Tōru Nakai)
Faculty of Education, Chiba University

1　はじめに

[8] において、評価と関連する状態に関する状態をもとに、支出を決定する逐次決定問題を扱った。また、状態は支出することによって変えることが出来た。ここでは、[8] などで扱った問題を費用最小化問題に応用することを考える。

いま、自動車や電化製品などに関して問題が生じたとき、問題点が大きくなければメーカーは個別的に対応することが可能であるが、問題が大きく部品の取り替える必要性が生ずるような場合には、リコールを行うことになる。そこで、製品の状態を $(0,\infty)$ によって表し、状態を表す値 s が大きくなれば製品の抱える問題が大きくなるとする。また、考える製品に対するクレームの大きさは、状態に応じて異なり、非負の確率変数 X_s で表され、この確率変数の値を観測し、この値をもとに決定を行い、計画期間内で費用を最小化する問題を解析する。

2　逐次決定問題

状態空間を $(0,\infty)$ とし、状態を表す値 s が大きくなれば状態が悪くなるとする。また、状態 $s\in(0,\infty)$ に対して、非負の確率変数 X_s を仮定する。これらの確率変数は $E[X_s]<\infty$ であり、s に関して確率的に増加とする。すなわち、s の値が大きくなればなるほどクレームの大きさも大きくなると考える。また、状態に関して不完備情報の場合には、これらの確率変数を観測することを情報過程とし、この値を観測することによって状態に関する情報を得る。さらに、x を観測したとき x をクレームの大きさと考え、費用 $c(x)$ を支払ってクレームに対応する。いっぽう、リコールを行えば取り替えて問題を解決することができ、その費用を $C(s)$ とする。$u(s)$ を最後の期の状態が s のときの終端費用とし、$c(x)$ は x に関して非負の非減少関数とする。

状態は推移法則を $\boldsymbol{P}=(p_s(t))_{s,t\in(0,\infty)}$ とするマルコフ過程にしたがって推移し、確率変数 X_s とは独立とする。さらに、推移法則には、TP_2 を仮定する (定義 1)。

定義 1 推移法則 $\boldsymbol{P}=(p_s(t))_{s,t\in(0,\infty)}$ は、$s\le t$ および $u\le v$ となる任意の s,t,u と v に対して $(s,t,u,v\in(0,\infty))$、$\begin{vmatrix} p_s(u) & p_s(v) \\ p_t(u) & p_t(v) \end{vmatrix}\ge 0$ となる。

3　部分観測可能なマルコフ過程

状態空間が $(0,\infty)$ のマルコフ過程で、推移法則を $\boldsymbol{P}=(p_s(t))_{s,t\in(0,\infty)}$ とする。このとき、このマルコフ過程の状態を直接知ることが出来ない不完備情報のマルコフ過

程を考える。すなわち、このマルコフ過程の状態に関する情報は、状態空間 $(0,\infty)$ 上の確率分布 $\boldsymbol{\mu}$ で表し、$\mathcal{S}$ を状態に関する情報全体の集合とすれば、

$$\mathcal{S}=\left\{\boldsymbol{\mu}=(\mu(s))_{s\in(0,\infty)}\,\middle|\,\int_0^\infty \mu(s)ds=1,\mu(s)\geq 0\,(s\in(0,\infty))\right\}$$

となる。

$\mathcal{S}$ に含まれる情報のあいだに、定義 1 と同じように、TP_2 を用いて半順序を定義する。この問題では、s が大きくなれば状態は悪くなるので、$\boldsymbol{\mu}$ が $\boldsymbol{\nu}$ よりこの半順序の意味で大きいとき、状態に関する情報は悪い情報を多く含むことになる。

いっぽう、$\boldsymbol{p}_s=(p_s(u))$ および $\boldsymbol{p}_{s'}=(p_{s'}(u))$ とおけば、$\boldsymbol{P}$ が定義 1 を満たすことから、任意の $s,s'\,(s\leq s',\ s,s'\in(0,\infty))$ に対して、$\boldsymbol{p}_{s'}\succeq\boldsymbol{p}_s$ となる。このとき、これらの仮定のもとで、Kijima and Ohnishi[2] などからつぎの性質が成り立つ。

補題 1 $\boldsymbol{\mu}\succeq\boldsymbol{\nu}$ ならば $(\boldsymbol{\mu},\boldsymbol{\nu}\in\mathcal{S})$、$x$ に関する非減少な非負関数 $h(x)$ に対して、$\int_0^\infty h(x)dF_{\boldsymbol{\mu}}(x)\geq\int_0^\infty h(x)dF_{\boldsymbol{\nu}}(x)$ となる。

3.1 部分観測可能なマルコフ過程と情報

つぎに、それぞれの状態 s に対して、状態に依存する確率変数 X_s を情報プロセスとする。すなわち、それぞれの状態に関する情報を確率変数 X_s を通して得ることができる情報システムあるいは観測過程を考える。さらに、状態は直接には観測できず、状態に依存する確率変数 X_s を通じて情報が得られ $(s\in(0,\infty))$、学習プロセスはベイズ学習にしたがって解析することから、仮定 1 を設ける。状態 s に対して、確率変数 X_s は絶対連続で、密度関数 $f_s(x)$ を持つとする $(s\in(0,\infty))$。この仮定は、Nakai [6] にしたがって一般化でき、多段決定問題へ応用できる (Nakai [3, 4, 5] など)。

仮定 1 確率変数 $\{X_s\}_{s\in(0,\infty)}$ に対して、$s\leq s'$ ならば、$X_{s'}\succeq X_s$ である $(s,s'\in(0,\infty))$。すなわち、X_s は s に関して尤度比の意味で増加する。

仮定 1 より、確率変数 X_s は s の値が大きくなるにしたがって、大きな値をとるようになり、s が大きくなるにしたがって悪くなり、それとともにクレームの大きさが大きくなる。また、推移法則に関する仮定から、現在の状態から、より悪い状態に推移する確率は、現在の状態が悪くなるにしたがって増加する。すなわち、それぞれの状態を表す s が大きくなれば、より悪い状態に推移する確率は大きくなるのである。

すなわち、この確率変数を観測することによって、状態に関して学習を行う。事前情報を $\boldsymbol{\mu}$ とするとき

1. 確率変数 $\{X_s\}_{s\in(0,\infty)}$ を観測する
2. 観測値をもとにベイズの定理により情報を改良し、$\boldsymbol{\mu}^x=(\mu^x(s))_{s\in(0,\infty)}$ とする
3. この過程は、推移法則 $(p_{s(x)}(t))_{0<s<\infty}$ に従って状態が推移し、推移した時点での状態に関する事後情報は $\overline{\boldsymbol{\mu}^x}=(\overline{\mu^x(s)})_{s\in(0,\infty)}$ となる

このとき、集合値関数 $\boldsymbol{h}(x,s)$ に対して、半順序 $\succeq$ を定義すれば、事前情報 $\boldsymbol{\mu}$ と事後情報 $\overline{\boldsymbol{\mu}(x)}$ のあいだには、マルコフ過程の推移法則に関する仮定と仮定 1 のもとで、つぎの基本的な性質が成り立つ (Nakai [6, 7] など)。

補題 2 $\boldsymbol{\mu} \succ \boldsymbol{\nu}$ ならば、任意の y に対して、$\boldsymbol{\mu}^x \succ \boldsymbol{\nu}^x$ および $\overline{\boldsymbol{\mu}^x} \succ \overline{\boldsymbol{\nu}^x}$ である。任意の $\boldsymbol{\mu}$ に対して、$\boldsymbol{\mu}^x$ と $\overline{\boldsymbol{\mu}^x}$ は x に関する増加関数である。

補題 2 から、事前情報 $\boldsymbol{\mu}$ における順序関係は、$\boldsymbol{\mu}^x$ と事後情報 $\overline{\boldsymbol{\mu}^x}$ に対して保たれる。さらに、同じ事前情報 $\boldsymbol{\mu}$ であれば、観測した値 x が大きくなれば、事後情報 $\overline{\boldsymbol{\mu}^x}$ もまたよくなる。

4 不完備情報の多段決定問題

部分観測可能なマルコフ過程で考える。すなわち、クレームの大きさによって状態に関する情報を得て決定を行う場合である。

4.1 決定が推移に影響する多段決定問題

クレームの大きさを見て部品を取り替えてリコールするとき、[8] で考えたように、どの程度まで取り替えるかといったように、クレームの大きさによって対応を変化させることが出来る場合を考える。すなわち状態 s を、観測値 x によって、変化させることができると考える。部分的にでもクレームに対応することで、状態が良くなり、その変化の度合いは決定に依存する場合である。

このとき、$w_n(s)$ を計画期間が n で状態が s のとき、最適に振る舞って得られる総期待費用とし、$w_n(s|x)$ を計画期間が n で状態が s のとき x を観測し、最適に振る舞って得られる総期待費用すれば、最適性の原理より、

$$w_n(s) = E_{X_s}[w_n(s|X_s)]$$

$$w_n(s|x) = c(x) + \min_{0\le\alpha\le1}\{C(\alpha) + w_{n-1}(\alpha s), w_{n-1}(s)\} \tag{1}$$

とする。ただし、$w_1(s|x) = c(x) + \min_{0\le\alpha\le1}\{C(\alpha) + u(\alpha s)\}$ とする。このとき、観測値 (クレームの大きさ) によって、クレームにどの様に対応できるかは、現在の状態 s の大きさに関わりなく、等倍率で状態を良くできるとし、そのための費用は絶対量ではなく倍率で定まると考える。すなわち、状態が s のとき、この状態は s の α 倍にすることができ ($0 < \alpha \le 1$)、状態を α 倍だけよくするための費用を $C(\alpha) = -\log\alpha$ とする。この $C(\alpha)$ は状態を α 倍だけよくするための費用だから、α に関して減少関数である。

$u(s)$ が s に関する増加関数だから、$w_1(s|x)$ も s に関する増加関数である。さらに帰納法により、$w_{n-1}(s)$ が s の増加関数で、$w_{n-1}(\alpha s)$ は α の増加関数だから、$w_n(s|x)$ も s に関する増加関数である。したがって、$w_n(s)$ も s に関する増加関数となる。また、$w_{n-1}(\alpha s)$ も、α の増加関数である。さらに、$\alpha = 1$ のときは、$v_{n-1}(\alpha s) = v_{n-1}(s)$ であり、$\alpha = 0$ のときは、$v_{n-1}(\alpha s) = v_{n-1}(0)$ である。

つぎに、状態がマルコフ過程にしたがって推移する場合を考える。いま、$\bar{w}_n(s)$ を計画期間が n で状態が s のとき、最適に振る舞って得られる総期待費用とし、$\bar{w}_n(s|x)$ を計画期間が n で状態が s のとき x を観測し、最適に振る舞って得られる総期待費用すれば、最適性の原理より、

$$\bar{w}_n(s) = E_{X_s}[\bar{w}_n(s|X_s)]$$

$$\bar{w}_n(s|x) = c(x) + \min_{0\le\alpha\le 1}\{C(\alpha) + \int p_{\alpha s}(t)\bar{w}_{n-1}(t)dt, \int p_s(t)\bar{w}_{n-1}(t)dt\} \tag{2}$$

とする。ただし、$\bar{w}_1(s|x) = c(x) + \min_{0\le\alpha\le 1}\{C(\alpha) + u(\alpha s)\}$ とする。このとき、つぎの性質が成り立つ。

補題 3 $\bar{w}_n(s)$ は s の増加関数であり、$\bar{w}_n(s|x)$ は s と x の増加関数である。

非負関数 $u(s), C(\alpha)$ は $0 < \lambda < 1$ に対して、$u(\widehat{s}^{\lambda}\overline{s}^{1-\lambda}) \le \lambda u(\widehat{s}) + (1-\lambda)u(\overline{s})$ を仮定する。

補題 4 $\widehat{s} < \overline{s}$ のとき、非負関数 $u(s)$ が $0 < \lambda < 1$ に対して、$u(\widehat{s}^{\lambda}\overline{s}^{1-\lambda}) \le \lambda u(\widehat{s}) + (1-\lambda)u(\overline{s})$ ならば、$\widehat{s} < \overline{s}, \widehat{s}' < \overline{s}'$ となる $\widehat{s}, \overline{s}, \widehat{s}', \overline{s}'$ に対して、$\dfrac{u(\widehat{s}) - u(\overline{s})}{\log\widehat{s} - \log\overline{s}} \le \dfrac{u(\widehat{s}') - u(\overline{s}')}{\log\widehat{s}' - \log\overline{s}'}$ である。

補題 5

$$w_n(s) = E_{X_s}[w_n(s|X_s)]$$

$$w_n(s|x) = c(x) + \min_{0\le\alpha\le 1}\{C(\alpha) + w_{n-1}(\alpha s)\}, \tag{3}$$

ただし $w_1(s|x) = c(x) + \min_{0\le\alpha\le 1}\{C(\alpha) + u(\alpha s)\}$ とするとき、$w_n(s), w_n(s|x)$ は、補題 *4* の性質を満たす。

補題 6 $\alpha_n(s)$ は s に関して減少する。

補題 7 $C(\alpha)$ が凸関数であり、$u(s)$ が *convex* 非減少関数であれば、$f(s) = \min_{0\le\alpha\le 1}\{C(\alpha) + u(\alpha s)\}$ もまた s に関する凸関数である。

4.2 Gradually Condition

[8] において、状態空間を $(-\infty, \infty)$ のとき、不完備情報のマルコフ過程での最適決定問題を考えるための条件を考えた。[8] で考えた支出モデルでは、決定がつぎの期の状態に影響することからもこれらの条件が必要であった。このなかで、状態が s のとき決定 x をとれば、状態は $s(x) = s + d(x)$ となると仮定した。このとき、$d(x)$ は、$d(0) = 0$ で、x に関する増加関数である。このとき、

$\boldsymbol{\mu}$: 事前情報

$\boldsymbol{\mu}_y = (\mu_y(s))$: 事前情報が $\boldsymbol{\mu}$ のとき、決定 y を取ったあとでの状態空間上の分布 $\overline{\mu_y}(s) = \int_{-\infty}^{\infty}\mu_y(t)p_t(s)dt = \int_{-\infty}^{\infty}\mu(t)p_{t(y)}(s)dt$

$\overline{\boldsymbol{\mu}_y} = (\overline{\mu_y}(s))$: 事前情報が $\boldsymbol{\mu}$ のとき、決定 y を取ったあと、推移法則にしたがって状態が推移したあとでの状態空間上の分布

とする。ここで、$s(0) = s$ だから、$\overline{\boldsymbol{\mu}} = \int_0^{\infty}\mu(s)p_s(t)ds = \boldsymbol{\mu}_0$ である。

さらに、状態の推移、学習、決定と事後情報との関係を見るため、つぎの性質と仮定を置いた。

定義 2 集合 $\mathcal{S}$ に含まれる確率分布 $\boldsymbol{\mu}$ が「$s < t, s' < t'$ と $s - s' = t - t' = c < 0$ を満たす任意の $s < s'$, $t \leq t'$ に対して、$\dfrac{\mu(s)}{\mu(s')} \geq \dfrac{\mu(t)}{\mu(t')}$」の性質を満たすとき、この $\boldsymbol{\mu}$ は *gradually condition* を満足するということにする。

補題 8 集合 $\mathcal{S}$ に含まれる確率分布 $\boldsymbol{\mu}$ が *gradually condition* を満足するとき、$x > x'$ ならば、$\boldsymbol{\mu}^x \succeq \boldsymbol{\mu}^{x'}$ である。

つぎの性質を導くため、推移法則に関してつぎの仮定をおく。

仮定 2 任意の $s < s'$, $t \leq t'$ および $u < v$ となる s, s', t, t', u, v に対して $p_u(s)p_v(t') - p_u(t)p_v(s') \geq p_v(s)p_u(t') - p_v(t)p_u(s')$ とする。

補題 9 集合 $\mathcal{S}$ に含まれる確率分布 $\boldsymbol{\mu}$ が *gradually condition* を満足するならば、$\overline{\boldsymbol{\mu}}$ もまた *gradually condition* を満足する。任意の x と決定 y に対して $\overline{\boldsymbol{\mu}^x}$ と $\overline{\boldsymbol{\mu}_y}$ も *gradually condition* を満足する。

4.3 単調性

事前情報を $\boldsymbol{\mu}$ とするとき、事後情報をつぎのように定義する。

$\boldsymbol{\mu}$: 事前情報、状態空間上の確率分布

$\overline{\boldsymbol{\mu}}$: 事前情報が $\boldsymbol{\mu}$ のとき、推移法則にしたがって状態が推移したあとでの状態空間上の分布

$\boldsymbol{\mu}_y$: 事前情報が $\boldsymbol{\mu}$ のとき、決定 y を取ったあとでの状態空間上の分布

$\boldsymbol{\mu}^x$: 事前情報が $\boldsymbol{\mu}$ のとき、観測値 x をもとにしてベイズの定理にしたがい改良した分布

決定と推移および学習の順序はつぎのように考える。

1. 情報過程から観測値 x を得る
2. ベイズの定理にしたがって、情報を $\boldsymbol{\mu}^x$ と改良する
3. 制約条件の中で、決定 y を取る
4. 推移法則 $\boldsymbol{P}$ にしたがって、この確率過程が推移する
5. つぎの時点における状態に関する情報は $\overline{\boldsymbol{\mu}_y^x} = (\overline{\mu_y^x}(s))$ である

補題 10 状態全体の集合 $\mathcal{S}$ に含まれる確率分布 $\boldsymbol{\mu}$ と $\boldsymbol{\nu}$ が *gradually condition* を満足するとき、$\boldsymbol{\mu} \succeq \boldsymbol{\nu}$ ならば、任意の $x(\geq 0)$ に対して $\overline{\boldsymbol{\mu}^x} \succeq \overline{\boldsymbol{\nu}^x}$ である。

簡単な計算から、任意の x に対して推移法則 $(p_{s(x))}(t))_{0 \leq s \leq 1}$ が TP_2 であるから、これまでに議論してきた仮定の下で、つぎの性質が成り立つ。

補題 11 状態全体の集合 $\mathcal{S}$ に含まれる確率分布 $\boldsymbol{\mu}$ と $\boldsymbol{\nu}$ が *gradually condition* を満足するとする。$x > x'$ ならば $\overline{\boldsymbol{\mu}(x)} \succeq \overline{\boldsymbol{\mu}(x')}$ である。$y > y'$ ならば $\boldsymbol{\mu}_y \succeq \boldsymbol{\mu}_{y'}$ であり、$\overline{\boldsymbol{\mu}_y^x} \succeq \overline{\boldsymbol{\mu}_{y'}^x}$ である。$\boldsymbol{\mu} \succeq \boldsymbol{\nu}$ ならば、任意の $y(\geq 0)$ に対して $\boldsymbol{\mu}_y \succeq \boldsymbol{\nu}_y$ であり、$\overline{\boldsymbol{\mu}_y} \succeq \overline{\boldsymbol{\nu}_y}$ および $\overline{\boldsymbol{\mu}_y^x} \succeq \overline{\boldsymbol{\nu}_y^x}$ である。$x > x'$ ならば、$\overline{\boldsymbol{\mu}_y^x} \succeq \overline{\boldsymbol{\mu}_y^{x'}}$ である。

4.4 対数正規分布

確率変数 Y を正規分布 $Y \sim N(\mu, \sigma^2)$ とするとき、$X := e^Y$ で定義される確率変数を対数正規分布といい、$y > 0$ のとき、事象 $\{X \le x\}$ と事象 $\{Y \le \log x\}$ は等しいので、X の密度関数 $f_X(x)$ は $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}}$ である。いま、正規分布 $N(\mu, \sigma^2)$ の密度関数を $\phi(x)$ とすれば、$f_X(x) = \phi(\log x)$ だから、$f_X(\alpha x) = \phi(\log \alpha x) = \phi(\log \alpha + \log x)$ となる。

[8] で扱った、評価を考慮した支出モデルでは、状態が s のとき、決定 x をとれば、状態を $s(x) = s + d(x)$ となる場合に、不完備情報のマルコフ過程での多段決定問題の性質を、gradually condition の性質を仮定して考えた。ここでは、状態が s のとき、決定 α をとれば、状態を αs と仮定した。すなわち、$s(\alpha) = \alpha s$ と考えればよい。よって、状態全体の集合 $\mathcal{S}$ に含まれる確率分布 $\boldsymbol{\mu}$ が

> $s < t, s' < t'$ と $\frac{s}{s'} = \frac{t}{t'} = \alpha < 0$ を満たす任意の $0 < s < s', 0 < t < t'$ に対して、$\frac{\mu(s)}{\mu(s')} \ge \frac{\mu(t)}{\mu(t')}$ となる

の性質を満たすときを考える。上記の性質から、集合 $\mathcal{S}$ に含まれる確率分布 $\boldsymbol{\mu}$ として対数正規分布を考えれば、この条件を満足するので、この場合を考える。

正規分布 $N(\mu, \sigma^2)$ の密度関数 $\phi(x|\mu, \sigma^2)$ は gradually condition を満たすので、情報プロセスを表す確率変数 X_s が対数正規分布であれば、$s < t, s' < t'$ となる $0 < s < s'$ と $0 < t < t'$ で $\frac{s}{s'} = \frac{t}{t'} = \alpha < 0$ であれば、$\frac{\mu(s)}{\mu(s')} \ge \frac{\mu(t)}{\mu(t')}$ となる。

正規分布 $N(\mu, \sigma^2)$ は TP_2 となる。いっぽう $f_X(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}}$ が対数正規分布の密度関数であれば、$f_X(x|\mu, \sigma^2) = \frac{\phi(\log x|\mu, \sigma^2)}{x}$ だから、σ^2 が μ に関する単調関数ならば、これらの確率変数は TP_2 となる。

観測できない状態に関する情報は、状態空間上の確率分布で表されているとしたが、ここではとくに $\boldsymbol{\mu}$ が状態空間 $(0, \infty)$ 上の対数正規分布によって表されているとする。このとき $\boldsymbol{\mu}$ に対し、事後情報をつぎのように定義する。$\mathcal{S}$ に含まれる事前情報 $\boldsymbol{\mu}$ が対数正規分布にしたがうとする。

- $\boldsymbol{\mu}$: 事前情報
- $\boldsymbol{\mu}^x$: トラブルの大きさ x を用いて、ベイズの定理にしたがって改良した分布
- $\boldsymbol{\mu}_\alpha$: 事前情報が $\boldsymbol{\mu}$ のとき、決定 α を取ったあとでの分布
- $\overline{\boldsymbol{\mu}^x_\alpha}$: 事前情報が $\boldsymbol{\mu}^x$ のとき、決定 α を取り、そのあと推移法則 $\boldsymbol{P}$ にしたがって状態が推移したあとでの分布

ここでは、事前情報が $\boldsymbol{\mu}$ のとき、はじめにトラブルの大きさ x を情報として観測し、ベイズの定理にしたがって情報を $\boldsymbol{\mu}^x$ と改良する。そのあと、decision-maker が決定 α をとり、, 状態に関する新しい情報を ${\boldsymbol{\mu}^x}_\alpha$ とする。さいごに、推移法則 $(p_s(t))_{0<s\le\infty}$ にしたがって、この過程の状態は推移し、状態に関する新野情報は $\overline{{\boldsymbol{\mu}^x}_\alpha}$ となる。

推移法則 $(p_s(t))_{0<s\leq\infty}$ について、任意の状態 $0<s\leq\infty$ に対して $(p_s(t))$ を状態空間上の対数正規分布とする。また、確率変数 X_s を正規分布にしたがうと仮定したとき、Nakai [8] で得られた単調性からつぎの性質が導かれる。

補題 12 $\boldsymbol{\mu}\succeq\boldsymbol{\nu}$ ならば、任意の観測値 x に対して、$\boldsymbol{\mu}^x\succeq\boldsymbol{\nu}^x$ である。$x>x'$ ならば、任意の事前情報 $\boldsymbol{\mu}$ に対して、$\boldsymbol{\mu}^x\succeq\boldsymbol{\mu}^{x'}$ である。

補題 13 $\alpha>\beta$ ならば、任意の事前情報 $\boldsymbol{\mu}$ に対して、$\boldsymbol{\mu}_\alpha\succeq\boldsymbol{\mu}_\beta$ である。

補題 14 $\alpha>\beta$ ならば、任意の観測値 x と事前情報 $\boldsymbol{\mu}$ に対して、$\boldsymbol{\mu}^x_\alpha\succeq\boldsymbol{\mu}^x_\beta$ である。$\boldsymbol{\mu}\succeq\boldsymbol{\nu}$ ならば、任意の観測値 x と決定 α に対して、$\boldsymbol{\mu}^x_\alpha\succeq\boldsymbol{\nu}^x_\alpha$ である。$x>x'$ ならば、任意の事前情報 $\boldsymbol{\mu}$ と決定 α に対して、$\boldsymbol{\mu}^x_\alpha\succeq\boldsymbol{\mu}^{x'}_\alpha$ である。

補題 15 $\boldsymbol{\mu}\succeq\boldsymbol{\nu}$ ならば、任意の観測値 x と決定 α に対して、$\overline{\boldsymbol{\mu}_\alpha}\succeq\overline{\boldsymbol{\nu}_\alpha}$ および $\overline{\boldsymbol{\mu}^x_\alpha}\succeq\overline{\boldsymbol{\nu}^x_\alpha}$ である。$x>x'$ ならば、任意の事前情報 $\boldsymbol{\mu}$ と決定 α に対して、$\overline{\boldsymbol{\mu}^x_\alpha}\succeq\overline{\boldsymbol{\mu}^{x'}_\alpha}$ である。

補題 16 $\alpha>\beta$ ならば、任意の観測値 x と事前情報 $\boldsymbol{\mu}$ に対して、$\overline{\boldsymbol{\mu}^x_\alpha}\succeq\overline{\boldsymbol{\mu}^x_\beta}$ である。

4.5 不完備情報の多段決定モデル

状態がマルコフ過程にしたがって推移し、その状態を直接知ることができず、クレームの大きさによって状態に関する情報を得る場合の逐次決定問題を考えることにしよう。クレームの大きさを知ることを、状態に関する情報を得る情報プロセスと考える。したがって、このモデルは、3 節の部分観測可能なマルコフ過程での逐次決定問題として定式化できる。

このような部分観測可能なマルコフ過程での逐次決定問題において、観測できない状態に関する情報は、状態空間上の確率分布として表され、前節で考えた性質を持つものとする。このとき、クレームの大きさを観測値とし、この値をもとにベイズの定理にしたがって学習を行う。また、3 節の部分観測可能なマルコフ過程においては、それぞれの状態 s $(s\in(0,\infty))$ に対して、クレームの大きさを表す確率変数 X_s を観測過程と考え、この値を観測することが情報プロセスとなる。

1. トラブルの大きさ x を状態に関する情報として観測する
2. この観測値をもとに、ベイズの定理にしたがって情報を $\boldsymbol{\mu}^x$ と改良する
3. 状態に関する情報が $\boldsymbol{\mu}^x$ のとき決定 α を取る。決定を取ったあとでの状態に関する情報は $\boldsymbol{\mu}^x{}_\alpha$ である。
4. 確率過程は 1 期進む
5. この確率過程は推移法則 $(p_s(t))_{0<s\leq\infty}$ にしたがって状態が推移し、推移した状態に関する情報は $\overline{\boldsymbol{\mu}^x{}_\alpha}$ となる
6. このとき、最適政策にしたがったときの総期待損失が $\widetilde{w}_{n-1}(\overline{\boldsymbol{\mu}^x{}_\alpha})$ である。

いま、状態に関する事前情報を $\boldsymbol{\mu}$ とし、計画期間が n のとき、最適政策にしたがって得られる総期待費用を $\widetilde{w}_n(\boldsymbol{\mu})$ とする。このとき、最適性の原理より、つぎの再帰方程式が得られる。

$$\widetilde{w}_n(\boldsymbol{\mu}) \quad = \quad \int_0^\infty \widetilde{w}_n(\boldsymbol{\mu}|x)d\boldsymbol{\mu}^x$$

$$\widetilde{w}_n(\boldsymbol{\mu}|x) \quad = \quad c(x) + \min_{0 \leq \alpha \leq 1} \{C(\alpha) + \widetilde{w}_{n-1}(\widetilde{\boldsymbol{\mu}^x{}_\alpha})\} \tag{4}$$

ここで、$\widetilde{w}_0(\boldsymbol{\mu}) = \int_0^1 u(t)d\mu(t)$ とする。$\alpha = 0,$ のときは、$\widetilde{w}_{n-1}(\widetilde{\boldsymbol{\mu}^x{}_\alpha}) \equiv 0$ とする。(4) 式において、$\boldsymbol{\mu}^x$ は情報過程から得られた観測値 x をもとに改良した事後情報とする。事前情報と事後情報に関する単調性から、これまでに考えた条件の下で $\boldsymbol{\mu} \succ \boldsymbol{\nu}$ ならば、任意の決定 α と観測値 x に対して、$\widetilde{\boldsymbol{\mu}^x{}_\alpha} \succeq \widetilde{\boldsymbol{\nu}^x{}_\alpha}$ である。よって、n に関する帰納法を用いれば、つぎの性質が得られる。.

性質 1 状態全体の集合 $\mathcal{S}$ に含まれる確率分布 $\boldsymbol{\mu}$ と $\boldsymbol{\nu}$ が前節の条件を満足するとき、$\boldsymbol{\mu} \succeq \boldsymbol{\nu}$ ならば、$\overline{w}_n(\boldsymbol{\mu}) \geq \overline{w}_n(\boldsymbol{\nu})$ である。

参考文献

[1] F. De Vylder, Duality Theorem for Bounds in Integrals with Applications to Stop Loss Premiums, *Scandinavian Actuarial Journal*, 129–147, (1983).

[2] M. Kijima and M. Ohnishi, Stochastic Orders and Their Applications in Financial Optimization, *Mathematical Methods of Operations Research*, **50**, 351–372, (1999).

[3] T. Nakai, A Sequential Stochastic Assignment Problem in a Partially Observable Markov process, *Mathematics of Operations Research*, **11**, 230–240, (1986).

[4] T. Nakai, An Optimal Selection Problem on a Partially Observable Markov process, In *Stochastic Modelling in Innovative Manufacturing*, Lecture Notes in Economics and Mathematical Systems **445**, (Eds. A. H. Christer, S. Osaki and L. C. Thomas), pp. 140–154, Springer-Verlag, Berlin, (1996).

[5] T. Nakai, An Optimal Assignment Problem for Multiple Objects per Period – Case of a Partially Observable Markov process, *Bulletin of Informatics and Cybernetics*, **31**, 23–34, (1999).

[6] T. Nakai, A Generalization of Multivariate Total Positivity of Order Two with an Application to Bayesian Learning Procedure, *Journal of Information & Optimization Sciences*, **23**, 163–176, (2002).

[7] T. Nakai, A Sequential Expenditure Problem for Public Sector Based on the Outcome, *Recent Advances in Stochastic Operations Research* (Eds. T. Dohi, S. Osaki and K. Sawaki), World Scientific Publishing, 277–295, 2007.

[8] T. Nakai, A Sequential Decision Problem based on the Rate Depending on a Markov Process, *Recent Advances in Stochastic Operations Research 2* (Eds. T. Dohi, S. Osaki and K. Sawaki), World Scientific Publishing, 11–30, 2009.